Inferring Intent and Action from Gaze in Naturalistic Behavior : A Review

Lukander, Kristian

2017-10

Lukander , K , Toivanen , M & Puolamäki , K 2017 , ' Inferring Intent and Action from Gaze in Naturalistic Behavior : A Review ' , International Journal of Mobile Human Computer Interaction , vol. 9 , no. 4 , pp. 41-57 . https://doi.org/10.4018/IJMHCI.2017100104

http://hdl.handle.net/10138/234727
https://doi.org/10.4018/IJMHCI.2017100104

unspecified
publishedVersion

International Journal of Mobile Human Computer Interaction
Volume 9 • Issue 4 • October-December 2017

Inferring Intent and Action from Gaze in Naturalistic Behavior:
A Review

Kristian Lukander, Finnish Institute of Occupational Health, Helsinki, Finland

Miika Toivanen, Finnish Institute of Occupational Health, Helsinki, Finland & Department of Teacher Education, University of Helsinki, Helsinki, Finland

Kai Puolamäki, Finnish Institute of Occupational Health, Helsinki, Finland

ABSTRACT

We constantly move our gaze to gather acute visual information from our environment. Conversely, as originally shown by Yarbus in his seminal work, the elicited gaze patterns hold information over our changing attentional focus while performing a task. Recently, the proliferation of machine learning algorithms has allowed the research community to test the idea of inferring, or even predicting action and intent from gaze behaviour. The on-going miniaturization of gaze tracking technologies toward pervasive wearable solutions allows studying inference also in everyday activities outside research laboratories. This paper scopes the emerging field and reviews studies focusing on the inference of intent and action in naturalistic behaviour. While the task-specific nature of gaze behavior, and the variability in naturalistic setups present challenges, gaze-based inference holds a clear promise for machine-based understanding of human intent and future interactive solutions.

KEYWORDS

Eye Movements, Gaze Tracking, Inference, Intent Modeling, Scoping Study, Task Modeling

INTRODUCTION

Gaze tracking in psychological, cognitive, and user interaction studies has recently evolved toward mobile solutions, which enable direct assessment of users' visual attention in natural environments. The capability for reliably tracking users' locus of attention with wearable devices has developed quickly as the device manufacturers have miniaturized their technology to wearable eye-glass-like frames, with a number of open-source solutions adding their contribution to the variety[1]. Also, increases in signal processing power and recent developments in gaze tracking algorithms now enable complex tracking methods to operate in portable devices, even in real-time (Toivanen et al., 2017).

Human eye movements shift the focus of attention to gather visual information for action planning. Conversely, they can be used to provide information for inferring users' intentions and next actions. However, gaze behavior in natural, unstructured tasks is markedly complex. Models created in controlled laboratory environments do not often satisfactorily explain such natural gaze behavior. While laboratory studies in gaze tracking typically aim for isolating single components of

DOI: 10.4018/IJMHCI.2017100104

behavior to accurately model and study some part of the human visual system or cognition, natural gaze behavior involves a complex interplay of these cognitive processes. The modeling of these processes computationally is difficult, not least because of the unknowns involved: it is a challenge to construct an experimental setup with a known "ground truth" for training, e.g., a machine learning model. In addition, the methods and implementations of machine learning applied to gaze data are still often customized and fine-tuned for each task at hand. This results in a set of isolated, individual contributions to gaze-based inference which are slowly converging to a more generic understanding on gaze-action behavior.

The issue of inferring user action with mobile gaze tracking is highly multidisciplinary, requiring deep understanding of a variety of research fields. These include the functioning of human visual system, mathematical modeling, computer vision, machine learning, cognitive processes, user interaction, and psychology. Here, we review current advances in attempting to infer the cognitive task of users based on their gaze behavior.

BACKGROUND

Motivation

Work toward this paper started from organizing the workshop[2] on "Inferring user action with mobile gaze tracking" as part of the Mobile HCI 2016 conference in Florence, Italy (Toivanen et al., 2016). The objective of the workshop was to map out the developing field of task and intent recognition in natural gaze interaction. The round-up talk after the workshop forms the basis of this contribution.

Eye and Gaze

The human visual system constantly samples the environment through a spatial window, where – due to the distribution of photoreceptor cells on the retina – high acuity information can only be obtained from the central area of the fovea, spanning about 1.5 degrees of visual angle. While the percept we experience seems stable, we inspect the scene through a constant stream of rapid, ballistic eye movements, saccades, to acquire new features from within the visual field. The acquisition of information takes place in between saccades, when the eye stabilizes the retinal image during fixations and slow smooth pursuit movements.

Eye movements, and the resulting gaze paths are highly task and context-specific (Rothkopf et al., 2015): the duration of a fixation is correlated with the complexity of the task performed and the information observed, and the distribution and time-course of saccades across visual stimuli holds information on the task performed. The active nature of eye movements when performing a task makes gaze direction a good proxy of attentional focus and even the underlying internal cognitive and contextual state.

Vergence movements (convergent, independent movement of the eyes) can provide further evidence on the depth plane of visual focus in binocular viewing. Pupillometry – the study of changes in pupil size – has also proved to provide information on cognitive activity, but as these are masked by pupil reactivity to luminosity variations in the stimulus environment their application in naturalistic settings seems unlikely. Eyelid movements and blinks, however, provide a natural addition to the trackable features of visual activity, e.g., increasing concentration appears to reduce blink frequency (Wang et al., 2014). Sleepiness and shifts in vigilance have been shown to be reflected in blink duration, amplitude and eye closing times (Papadelis et al., 2007; Morris and Miller, 1996).

Tracking Methods

Tracking eye movements and gaze has grown to a rich methodology for tracking the oculomotor activity, attentional focus, and cognitive activity of a user or patient population. The two most firmly established eye tracking techniques at present are the electro-oculogram (EOG) and video-oculography (VOG).

EOG is generated from measurements of electrical activity associated with eye movements, quantified by recording from electrodes applied to the skin surface around the eye(s). EOG setups vary but are most often performed with four electrodes: a pair of horizontal electrodes at the outer canthi of the eyes, summing the movement of the electrical dipoles of the two eyes; and a pair of vertical electrodes, placed above and below one eye for tracking vertical eye movements and eyelid activity. EOG provides high temporal resolution (up to several kilohertz) and can even be used when the eyes are closed e.g., during sleep or at sleep onset. EOG however offers only a limited spatial resolution, has a drifting baseline, and exhibits high-frequency noise (Eggert, 2007). EOG is thus suitable for wearable devices to accurately track oculomotor parameters or contextual information, but less applicable for providing actual point-of-gaze.

Devices for VOG measurements are camera-based, tracking the movements of the eye via changes in visual features such as the pupil, iris, sclera, and reflections of light sources on the surface of the cornea. VOG naturally also provides pupillary measures and data on eyelid movement. In this review, we focus on mobile, natural settings, and thus concentrate on mobile eye tracking equipment. These vary in their capabilities, but provide better spatial resolution than EOG (around 0.5–2 degrees of visual angle), with frame rates, however, typically around 30–60 Hz depending on the system. Hence, VOG systems are better suited to tracking the point of gaze, examining gaze path and patterns, and utilizing event-based metrics, while some systems with higher frame rates can also deliver accurate oculomotor parameters.

Gaze Features

Eye trackers enable extraction of several parameters for each type of eye movement. For fixations, typical parameters are location, duration, frequency, and drift within fixations. For saccades, the usual parameters considered are frequency, duration, amplitude, average speed, and speed profiles. In addition to the eye, trackers can provide information on eyelid movement and blinks, and common parameters for these such as frequency, blink duration, and eyelid closing and opening times. More complex, derived parameters include dwell times (the sum of fixation times within a defined area of interest or object), gaze paths and patterns, the area covered, and the frequency, number of, and sequence of areas of interest visited in visual stimuli. Bulling et al. (2009) list 90 different parameters used for activity recognition demonstrating the breadth of possible information sources attainable.

Intent Modeling

In his formative work, Yarbus (see Tatler et al., 2010) examined gaze paths of subjects viewing a painting by Ilja Repin ("The unexpected visitor", 1884) under seven different cognitive tasks ranging from free examination to memory tasks, and estimating the social status and activity of the people depicted in the painting. Yarbus was the first to show that gaze patterns varied considerably under different instructions while observing the same visual stimulus – that gaze patterns can be used to reveal the observer's task. With the advent of machine learning approaches, there's a recent renaissance in studying the inverse question: can we infer a person's intentions, cognitive task, or attentional focus from observing their gaze behavior?

A considerable part of the inference work on gaze data has discussed the bottom-up approach: predicting fixation distributions based on the local saliency features of the presented (static) stimulus material. While saliency is likely to explain some of the attention-grabbing features of stimuli — especially in free-viewing conditions where task-related factors do not guide top-down processing of visual stimuli (Abolhassani & Clark, 2011) — it provides an overly simplistic answer to prediction of action and intent. Saliency models have recently been summarized by Borji & Itti (2014).

A step further from the bottom-up models, Oliva et al. (2003) integrate the overall "gist" of a scene for guiding visual search: contextual priming guides object search to the more probable location of a target object within the scene (tasked with looking for people in a street photograph, the attention is more likely to concentrate on the street-level, where people would be expected). More recently,

O'Connell and Walther (2014) suggest that salience-driven (exogenous) and content-driven, scene category based (endogenous) spatial attention can be dissociated and seem to influence attention in slightly different time frames. Image or scene salience has a stronger influence on gaze behavior in the initial cycles at around 600 ms of perception and in free-viewing situations without a task objective. Scene context and the "gist" kicks in at around 2000 ms, and after we have constructed a personal (3D) representation of the space around us through visual examination, salience becomes more likely to influence gaze at the very local level. Task and context related factors guide the gaze to different loci, or "prune the search tree" within the scene (e.g., Navalpakkam & Itti, 2002), and the salient features interact in fine-tuning the final location of fixations within narrow target windows.

Figure 1 presents a simple schematic of the propositions above. Here, we concentrate on studies inferring intent in active viewing circumstances in natural (and virtual) environments, and in approaches that include scene context — that is, approaching the objective of inference from the right-hand edge of the figure.

Machine Learning

Extracting useful information from gaze can be challenging, as the observed gaze pattern is the result of an extremely complicated process that includes the often-noisy measurement, the cognitive state, activity and current objectives of the user, and the (dynamic) features of the environment. While in addition the ground truth for any of these can almost never be perfectly known, gaze still contains useful information for modeling the task at hand. Unfortunately, no generally applicable methods that would work across conditions and circumstances are available.

The two main approaches to analyzing gaze are (1) making use of well-known statistics and models of cognitive processes, and (2) approaches based on machine learning. Simpler metrics, such as the (accumulated) gaze location can help, e.g., to distinguish whether a user has noticed a visual target, and statistics of gaze and stimulus features may be sufficient for some objectives. However, should the task require more complex understanding of user activity with difficult-to-model interactions, simple models cannot supply sufficient information, and we typically resort to machine learning to extract more intricate details of user behavior.

Successful application of machine learning requires knowledge about the underlying cognitive, physiological, and task-specific aspects. However, machine learning methods themselves are quite generic and independent of these details. Typically, the setup is that of supervised learning, in which the objective is to predict the class of eye movements, task type, or properties of target objects from gaze patterns and other contextual features. Machine learning methods often provide a "black box" solution, combining various sources of information, even in surprising and unintuitive ways, which may lead to unexpected results when applied outside their training context. The black box nature of the resulting solution impedes generalizability, and makes applying methods across real life conditions more difficult.

Figure 1. Modes of inference

Approach	bottom-up		top-down
Driver	exogenous	endogenous	
	stimulus structure	scene category/context "gist"	task
Objective	Infer fixation locations from stimulus structure		Infer cognitive task from gaze behavior

The machine learning methods typically used in modeling gaze can be roughly split into two main classes: First, general purpose high-performance classifiers such as support vector machines (SVM, Cristianini & Shawe-Taylor, 2000) or random forests (Breiman, 2001) can be used in the prediction task. Here, the choice of input features is critical: while the time series nature of the gaze need not to be directly modeled, this information is usually contained in the selected features. The second main approach is the direct application of time series methods such as Hidden Markov Models (HMM) or rule based algorithms. These may better capture the persistence of cognitive processing states, and therefore model human behavior more accurately. Literature on applying machine learning methods to gaze data ranges from pure natural/mobile context, e.g., in information retrieval (e.g., Granka, 2004; Puolamäki, 2005) to screening clinical populations (e.g., Tseng et al., 2013).

MATERIAL AND METHODS

The area of intent modeling in natural gaze tracking requires contributions from two very different fields: eye movement research and the related cognitive aspects, and the field of machine learning and pattern classification. We opted to perform study whose execution is detailed in Table 1. Our approach is motivated by the process for a scoping study by Levac et al. (2010), although we relax some of the more rigorous process steps due to the open nature of the application field and the available resources. Scoping studies are more routinely applied in the field of healthcare, and that aims to answer a broader need for scoping an area of literature to map key concepts and types of evidence available (Arksey & O'Malley, 2005), summarize and disseminate research findings, and/or identify gaps in the existing literature (Levac et al., 2010). We recognize that in a study like ours it is impossible to fit all relevant publications, alone due to limitations of the bibliographic databases and different terminologies used. Our purpose is instead to provide a representative sample of the contributions and thereby give an overview of the field.

Table 1. The procedure of our study, motivated by Levac et al. (2010)

1. Identify (broad) research question	We aim to answer the question: "How has gaze-based intent modelling been performed, in what (naturalistic) environments, and which approaches seem most promising?"
2. Identify and select relevant studies	The implemented search strategy aims for comprehensiveness and breadth while keeping the number of papers included within a controlled range. • We started off from with the papers presented in the workshop, and the work referenced in those papers • We then searched for additional sources from two databases: *Scopus* and *Web of Science* using Boolean permutations of keywords "(infer OR predict) AND (intent OR task) AND gaze" • We decided to exclude infant and animal research and medical and neurological conditions • As mobile tracking methods and machine learning methods have developed by leaps and bounds during the last decade, we limited the search further to the last ten years.
3. Study selection	This resulted in 181 (*Scopus*) + 255 (*WoS*) papers Representative papers were then selected based on their titles and abstracts, and after removing duplicates this resulted in 27 (WoS) + 17 (Scopus) papers After the final round of reading, 29 papers were included from the search, added with 2 from the workshop, and 4 papers known of by the authors outside the search result
4. Charting the data	The objectives and methods in the papers cover a large topic area, but the methods, success rates, equipment and features used were tabulated to the extent possible for a quick comparison chart.
5. Collate & summarize	Finally, an overview was provided. As the initial question does not have a definite answer, and the approach is exploratory, no numerical, or comparative analysis can be provided.

Generally, the inference literature can be characterized by a four-fold Table 2. The models applied in the papers can roughly be divided to bottom-up approaches, evaluating and predicting gaze behavior based on low-level features of the stimuli, or inferring task-specific behaviors based on top-down control. On the other hand, the bulk of the research has been done in controlled laboratory conditions, with simplified 2D generated/projected stimulus material, while some more recent works aim toward studying naturalistic behavior in real-world, or simulated, three-dimensional virtual environments.

RESULTS: INFERRING INTENT IN NATURAL ENVIRONMENTS

We aim to report a breadth-first view of the available literature, delivering a broad review of the application areas. The results of the initial literature search revealed that even with targeted keywords, the bulk of the papers deal with bottom-up, salience driven approaches. A general overview of the reviewed papers shows that the stimulus environments vary considerably, and as naturality dictates, sometimes even within studies. Also evident is considerable inter-individual variation in (gaze) behavior and responses but also in basic gaze tracking performance. The studies included in the literature search are summarized later in Table 3 (see Appendix).

For inference of (cognitive) task factors or task identity based on gaze behavior, the most popular application areas include car driving in both natural and simulator environments, path navigation, and variations of the inverse Yarbus process. In line with our expectations, several studies were performed using virtual/augmented reality as the stimulus environment. These afford a better way to control the stimuli, and deliver ground truth on, e.g., gaze targets, albeit limiting the naturality to an extent. There seems to be surprisingly few papers addressing real-world working life tasks such as installation work, industrial work, or routine work such as customer service while these could provide research with structured operational environments and relevant research applications.

Part of the papers approach recognition offline, from summary statistics, while fewer works attempt at inferring intent online applying running diagnostics. A few studies (Kit et al., 2016; Peng et al., 2015; Vrzakova & Bednarik, 2015) study the effective length of the prediction window: how long of a sample of task-related behavior is needed for inference, and how long before actual action can the presented solution deliver reliable predictions.

Car driving offers a semi-controlled "moving laboratory environment", where the subject stays relatively put in a well-controlled three-dimensional stimulus environment, while participating in a complex, dynamic task with continuous components (stay in lane), distractors, task objectives (navigation) etc. Peng et al. (2015) were able to predict online (accuracy 85.4% 1.5 s before initiation) when the driver was about to change lanes based on "visual search behavior" using a back-propagation neural network model. Another lane-changing study (Wen et al., 2015) used a hidden conditional random fields (HCRF) model combining gaze position and vehicle data, and showed that it was able to outperform SVM's and HMM's with a 99% recognition rate 0.5 s before lane change, and 85%

Table 2. Four-fold classification of inference papers. The analysis here will focus on the upper right-hand square.

Context \ Model	Bottom-Up	Top-Down
Natural (like) environments	Saliency in photographic stimuli (video, virtual reality)	Inferring user activity and intent through top-down understanding of gaze path activity in natural environments
Controlled 2D lab stimuli	Inference on probable fixation locations in free viewing or simple tasks for static stimuli based on feature salience etc.	Task-guided gaze activity with generated stimuli in static contexts

performance level 2.0 s in advance. Lethaus et al. (2013b) were able to predict lane change up to five seconds before the actual event. Johnson et al. (2014) approached task modeling in a dual-task driving scenario (adhere to a given speed requiring frequent gazes at the speedometer, follow a lead car) by decomposing visual behavior into individual task modules in order to model the distribution of gaze on task-relevant objects. Their softmax barrier model outperforms Itti & Koch (2001) saliency and central bias models in predicting fixations to task-relevant items, and they claim that model should be generalizable to other realms outside driving. Lethaus et al. (2013a) compared different machine learning algorithms to predict driver's intent and found out that artificial neural networks performed slightly better in their data than Bayesian networks and naive Bayesian classifier. Borji et al. (2012a) developed a Kernel Density Estimation method, combining both bottom-up and top-down influences in their modeling of driver's intent in a video game, and report outperforming the compared "state-of-the-art" methods by 15%.

The plethora of studies on different models of visual attention which can be roughly split into (1) bottom-up models such as saliency based approaches and (2) top-down models such as object-based theories. These have been studied e.g. in Borji et al. (2012b) and Borji & Tanner (2016). Mathe & Sminchisescu (2015) train saliency detectors based on actual fixation data and show that these can reliably predict human fixations in variable visual material.

In predicting user preference and attention allocation, Huang et al. (2015) succeeded in predicting which food ingredient a sandwich shop customer was about to ask for 1,8 s before the spoken request with a 76% accuracy by feeding simple gaze features to a SVM. Asteriadis et al. (2008) used gaze to infer user attentiveness reaching an 88% performance level, while Hamed et al. (2016) explored the problem of using Gaussian processes with gaze to assess users' preference between different keyword clouds, reaching a 63% classification accuracy in a binary classification task. Ajanki et al. (2011) integrated relevance estimation based on gaze intensity to an augmented reality headset.

While path navigation offers a seemingly simple, overlearned task, the resulting gaze behavior differs considerably from static scenes because of the complexities of dynamic interaction with the environment. t'Hart et al. (2012) provide summary statistics for gaze allocation in naturalistic path navigation. Rothkopf (2016) developed a codebook of gaze locations and modeled HMMs with varying numbers of latent variables able to generate gaze sequences comparable to actual human data in navigating an environment with targets and obstacles. Zank & Kunz (2016) succeeded in improving the prediction of user locomotion in virtual reality by utilizing gaze information.

Eye-hand coordination is a central activity in all our natural interactions. Carrasco & Clady (2010) combine an eye tracker with a camera attached to the user's hand, and report recognizing the reach to grab gesture with 80–90% probability. Vrzakova&Bednarik (2015) show that considering the "quiet eye" — the stable fixation just before action initiation, originally suggested by Vickers (1996) — can considerably increase predictive power, although within a considerably shortened time frame.

Information retrieval is another essential task within different contexts. Puolamäki et al. (2005), Puolamäki et al. (2008), Ajanki et al. (2009) are examples of using gaze trajectories in facilitating information retrieval by estimating the relevance of the text read by the user or of predicting the search terms relevant for the user. Liu et al. (2009) excelled in distinguishing novices and experts using HMMs while reading and manipulating concept maps. Voisin et al. (2013) were able to predict perceptual errors in reading mammography images using machine learning algorithms for fusing gaze and features from radiology images.

The (inverse) Yarbus process has recently received fair attention in the literature. This might be attributable to Greene et al. (2012) claiming that the task could not be performed, refuting Yarbus' original assertion. This was followed up by a set of work proving the opposite: Kanan Haji-Abolhassani & Clark (2013) successfully used HMMs to model the cognitive search process, Kanan et al. (2014) showed that with Greene's original data, prediction is possible using better algorithms.

Boisvert&Bruce (2015) applied random forests, and Borji & Itti (2015) used kNN with Boosting for good classification results. Vincent, 2012 modeled the different mechanisms for utilizing the past observations in predicting target's future location.

Work on general activity recognition was addressed in only a few works. Bulling et al. (2009) distilled 90 different features of eye movements measured using EOG, and used an SVM approach to obtain a 76% accuracy in recognizing user activity within five typical office activities with 70.5% recall over all subjects. Kit & Sullivan (2016) classified tasks between five different everyday activities from sandwich making to frisbee catching. Using HMMs for only time series data for saccadic direction and amplitude they reached an overall recognition performance of 36%, opposed to 20% chance level.

CONCLUSION AND FUTURE DIRECTIONS

Haji-Abolhassani and Clark (2014) labelled the extraction of intent from gaze pattern an "inverse Yarbus process", as Yarbus' original investigation was into the effect of instruction on gaze patterns. As confirmed by the current work as well as others, this presents a lucrative, yet demanding target for research offering numerous applications and considerable impact.

Simplified, the process of inferring intent from gaze walks through the following steps: record gaze data, identify and extract features, associate cognitive models and knowledge about human information processing (capacity, speed), train a machine learning model to recognize and classify states and behaviors, and apply this model in practice. Ultimately, the end product should do this in real time, without the wearer's intrusion or guidance, and deliver a reliable metric of things attended or actions intended, preferably proactively before the wearer has even initiated the associated motor action. Another objective is to broaden the bandwidth between man and machine through supplying reliable context recognition while performing tasks, applicable in use cases where the (devices within the) environment would "know" what the user wants without explicit communication. Yet another evident application is safety associated with human intention and activity in traffic, and demanding operational environments. Also, to escape the uncanny valley (Mori et al., 2012), future humanoid robots may well need to match humans in their natural understanding of other people's intentions, derived from minute behavioral hints.

Isolated gaze features or summary statistics of eye movements do not appear to elicit sufficient amounts of information to reliably identify the visual task performed (see also Haji-Abolhassani, 2014). However, this does not rule out the potential of other, more informative measures that consider the temporal dynamics of eye movements, or combine gaze-based information with other data regarding the target of operation or the operational environment. On the other hand, one of the pioneers of eye tracking research, Rayner (2009), warns that it might be hazardous to generalize eye movement metrics across even simple task types such as reading and visual search. As eye movement metrics are highly task- and subject-specific, movements in the real world can perhaps ultimately be understood only in the context of a particular task.

Often using some sort of persistent state models such as HMMs and Markov chains, fare better in deducing (sequences of) actions than time-agnostic classifiers (Griffiths et al., 2008). This is to be expected, as the human cognition shows similar persistence in performing a single task at a time, and it would seem that the strategy is to resort to rapid task-switching instead of "multitasking", even under time pressure.

Until the field successfully constructs standardized approaches and toolboxes for consistently and successfully inferring intent from gaze — possibly in combination with other psychophysical or environmental signals — the contributions are likely to stay isolated, task-dependent, and appropriate only within narrow application areas. As a large proportion of the existing studies have looked at

eye tracking in laboratory conditions, studying and applying gaze interaction and gaze-based user modelling in natural environments presents a substantial opportunity. However, individual-to-individual variability and the task-specific nature of eye movements should be carefully considered, if one is to deliver successful applications of eye-aware user interfaces and insights into the cognitive state of users.

ACKNOWLEDGMENT

We thank the participants of the workshop for their presentations and discussion. This work was funded by the Academy of Finland (decisions 286154 and 297856) and Tekes (Revolution of Knowledge Work Project).

REFERENCES

Abolhassani, A. H., & Clark, J. J. (2011). Visual Task Inference Using Hidden Markov Models. Proceedings of IJCAI (pp. 1678-1683).

Abolhassani, A. H., & Clark, J. J. (2011, June). Visual Task Inference Using Hidden Markov Models. Proceedings of IJCAI (pp. 1678-1683).

Ajanki, A., Billinghurst, M., Gamper, H., Järvenpää, T., Kandemir, M., Kaski, S., & Ruokolainen, T. et al. (2011). An augmented reality interface to contextual information. *Virtual Reality (Waltham Cross)*, *15*(2-3), 161–173. doi:10.1007/s10055-010-0183-5

Ajanki, A., Hardoon, D. R., Kaski, S., Puolamäki, K., & Shawe-Taylor, J. (2009). Can eyes reveal interest? Implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction*, *19*(4), 307–339. doi:10.1007/s11257-009-9066-4

Arksey, H., & OMalley, L. (2005). Scoping studies: Towards a Methodological Framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. doi:10.1080/1364557032000119616

Asteriadis, S., Karpouzis, K., & Kollias, S. (2008, September). A neuro-fuzzy approach to user attention recognition. *Proceedings of the International Conference on Artificial Neural Networks* (pp. 927-936). Springer Berlin Heidelberg. doi:10.1007/978-3-540-87536-9_95

Babcock, J. S., & Pelz, J. B. (2004) Building a lightweight eyetracking headgear. Proceedings of the ACM SIGCHI eye tracking research & applications symposium (pp. 109–114). doi:10.1145/968363.968386

Bernhard, M., Stavrakis, E., Hecher, M., & Wimmer, M. (2014). Gaze-to-object mapping during visual search in 3d virtual environments. *ACM Transactions on Applied Perception*, *11*(3), 14. doi:10.1145/2644812

Boisvert, J. F., & Bruce, N. D. (2016). Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features. *Neurocomputing*, *207*, 653–668. doi:10.1016/j.neucom.2016.05.047

Borji, A., & Itti, L. (2014). Defending Yarbus: Eye movements reveal observers task. *Journal of Vision (Charlottesville, Va.)*, *14*(3), 29–29. doi:10.1167/14.3.29 PMID:24665092

Borji, A., Sihite, D. N., & Itti, L. (2012, June). Probabilistic learning of task-specific visual attention. *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 470-477). IEEE. doi:10.1109/CVPR.2012.6247710

Borji, A., Sihite, D. N., & Itti, L. (2012, May). Modeling the influence of action on spatial attention in visual interactive environments. *Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 444-450). IEEE. doi:10.1109/ICRA.2012.6224551

Borji, A., & Tanner, J. (2016). Reconciling saliency and object center-bias hypotheses in explaining free-viewing fixations. *IEEE transactions on neural networks and learning systems*, *27*(6), 1214-1226.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi:10.1023/A:1010933404324

Bulling, A., Ward, J. A., Gellersen, H., & Tröster, G. (2009). Eye movement analysis for activity recognition. *Proceedings of the 11th international conference on Ubiquitous computing* (pp. 41-50). ACM.

Carrasco, M., & Clady, X. (2010, October). Prediction of user's grasping intentions based on eye-hand coordination. *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 4631-4637). IEEE. doi:10.1109/IROS.2010.5650024

Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines.

Eggert, T. (2007). Eye movement recordings: Methods. *Neuro-Ophthalmology (Aeolus Press)*, *40*, 15–34. PMID:17314477

George, A., & Routray, A. (2016). A score level fusion method for eye movement biometrics. *Pattern Recognition Letters*, *82*, 207–215. doi:10.1016/j.patrec.2015.11.020

Granka, L. A., Joachims, T., & Gay, G. (2004, July). Eye-tracking analysis of user behavior in WWW search. *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 478-479). ACM. doi:10.1145/1008992.1009079

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers task from eye movement patterns. *Vision Research*, *62*, 1–8. doi:10.1016/j.visres.2012.03.019 PMID:22487718

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian Models of Cognition. In R. Sun (Ed.), The Cambridge Handbook of Computational Psychology (pp. 59–100). Retrieved from https://www.cambridge.org/core/books/the-cambridge-handbook-of-computational-psychology/bayesian-models-of-cognition/58B8B762EEA8AB340140D9B98A83090B doi:10.1017/CBO9780511816772.006

Haji-Abolhassani, A., & Clark, J. J. (2013). A computational model for task inference in visual search. *Journal of Vision (Charlottesville, Va.)*, *13*(3), 29–29. doi:10.1167/13.3.29 PMID:24071637

Haji-Abolhassani, A., & Clark, J. J. (2014). An inverse Yarbus process: Predicting observers task from eye movement patterns. *Vision Research*, *103*, 127–142. doi:10.1016/j.visres.2014.08.014 PMID:25175112

Hamed, R., Poostchi, H., Peltonen, J., Laaksonen, J., & Kaski, S. (2016, December). Preliminary Studies on Personalized Preference Prediction from Gaze in Comparing Visualizations. *Proceedings of the International Symposium on Visual Computing* (pp. 576-585). Springer International Publishing.

Huang, C. M., Andrist, S., Sauppé, A., & Mutlu, B. (2015). Using gaze patterns to predict task intent in collaboration. *Frontiers in Psychology*, *6*, 1049. doi:10.3389/fpsyg.2015.01049 PMID:26257694

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience*, *2*(3), 194–203. doi:10.1038/35058500 PMID:11256080

Johnson, L., Sullivan, B., Hayhoe, M., & Ballard, D. (2014). Predicting human visuomotor behaviour in a driving task. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1636), 20130044. doi:10.1098/rstb.2013.0044 PMID:24395971

Kanan, C., Ray, N. A., Bseiso, D. N., Hsiao, J. H., & Cottrell, G. W. (2014, March). Predicting an observer's task using multi-fixation pattern analysis. *Proceedings of the symposium on eye tracking research and applications* (pp. 287-290). ACM. doi:10.1145/2578153.2578208

Kit, D., & Sullivan, B. (2016, September). Classifying mobile eye tracking data with hidden Markov models. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 1037-1040). ACM. doi:10.1145/2957265.2965014

Koehler, K., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict? *Journal of Vision (Charlottesville, Va.)*, *14*(3), 14–14. doi:10.1167/14.3.14 PMID:24618107

Lethaus, F., Baumann, M. R., Köster, F., & Lemmer, K. (2011, April). Using pattern recognition to predict driver intent. *Proceedings of the International Conference on Adaptive and Natural Computing Algorithms* (pp. 140-149). Springer Berlin Heidelberg. doi:10.1007/978-3-642-20282-7_15

Lethaus, F., Baumann, M. R., Köster, F., & Lemmer, K. (2013). A comparison of selected simple supervised learning algorithms to predict driver intent based on gaze data. *Neurocomputing*, *121*, 108–130. doi:10.1016/j.neucom.2013.04.035

Lethaus, F., Harris, R. M., Baumann, M. R., Köster, F., & Lemmer, K. (2013, April). Windows of driver gaze data: how early and how much for robust predictions of driver intent? *Proceedings of the International Conference on Adaptive and Natural Computing Algorithms* (pp. 446-455). Springer Berlin Heidelberg. doi:10.1007/978-3-642-37213-1_46

Levac, D., Colquhoun, H., & OBrien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, *5*(1), 69. doi:10.1186/1748-5908-5-69 PMID:20854677

Liu, Y., Hsueh, P. Y., Lai, J., Sangin, M., Nussli, M. A., & Dillenbourg, P. (2009, June). Who is the expert? Analyzing gaze data to predict expertise level in collaborative applications. *Proceedings of the IEEE International Conference on Multimedia and Expo ICME '09* (pp. 898-901). IEEE. doi:10.1109/ICME.2009.5202640

Mathe, S., & Sminchisescu, C. (2015). Actions in the eye: Dynamic gaze datasets and learnt saliency models for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(7), 1408–1424. doi:10.1109/TPAMI.2014.2366154 PMID:26352449

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. doi:10.1109/MRA.2012.2192811

Morris, T. L., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology*, *42*(3), 343–360. doi:10.1016/0301-0511(95)05166-X PMID:8652752

Navalpakkam, V., & Itti, L. (2002). A goal oriented attention guidance model. *Proceedings of the International Workshop on Biologically Motivated Computer Vision* (pp. 453-461). Springer Berlin Heidelberg. doi:10.1007/3-540-36181-2_45

OConnell, T. P., & Walther, D. B. (2015). Dissociation of salience-driven and content-driven spatial attention to scene category with predictive decoding of gaze patterns. *Journal of Vision (Charlottesville, Va.)*, *15*(5), 20–20. doi:10.1167/15.5.20 PMID:26067538

Oertel, C., Scherer, S., & Campbell, N. (2011). On the use of multimodal cues for the prediction of involvement in spontaneous conversation. In Interspeech (Vol. 2011, pp. 1541-1544).

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36. doi:10.1016/S0079-6123(06)55002-2 PMID:17027377

Papadelis, C., Chen, Z., Kourtidou-Papadeli, C., Bamidis, P. D., Chouvarda, I., Bekiaris, E., & Maglaveras, N. (2007). Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clinical Neurophysiology*, *118*(9), 1906–1922. doi:10.1016/j.clinph.2007.04.031 PMID:17652020

Peng, J., Guo, Y., Fu, R., Yuan, W., & Wang, C. (2015). Multi-parameter prediction of drivers lane-changing behaviour with neural network model. *Applied Ergonomics*, *50*, 207–217. doi:10.1016/j.apergo.2015.03.017 PMID:25959336

Puolamäki, K., Ajanki, A., & Kaski, S. (2008, July). Learning to learn implicit queries from gaze patterns. *Proceedings of the 25th international conference on Machine learning* (pp. 760-767). ACM. doi:10.1145/1390156.1390252

Puolamäki, K., Salojärvi, J., Savia, E., Simola, J., & Kaski, S. (2005, August). Combining eye movements and collaborative filtering for proactive information retrieval. *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 146-153). ACM. doi:10.1145/1076034.1076062

Puolamäki, K., Salojärvi, J., Savia, E., Simola, J., & Kaski, S. (2005, August). Combining eye movements and collaborative filtering for proactive information retrieval. *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 146-153). ACM. doi:10.1145/1076034.1076062

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. doi:10.1080/17470210902816461 PMID:19449261

Rothkopf, C. A. (2016, September). Minimal sequential gaze models for inferring walkers' tasks. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 1041-1044). ACM. doi:10.1145/2957265.2965015

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2015). Task and context determine where you look. *Journal of Vision (Charlottesville, Va.)*, *7*(14), 16. doi:10.1167/7.14.16 PMID:18217811

t'Hart, B., & Einhäuser, W. (2012). Mind the step: complementary effects of an implicit task on eye and head movements in real-life gaze allocation. *Experimental brain research*, *223*(2), 233-249.

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *i-Perception*, *1*(1), 7-27.

Toivanen, M., Lukander, K., Puolamäki, K. (2017) Probabilistic Approach to Robust Wearable Gaze Tracking. Manuscript submitted for publication.

Toivanen, M., Puolamäki, K., Lukander, K., Häkkinen, J., & Radun, J. (2016). Inferring user action with mobile gaze tracking. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 1026-1028). ACM. doi:10.1145/2957265.2965016

Tseng, P. H., Cameron, I. G., Pari, G., Reynolds, J. N., Munoz, D. P., & Itti, L. (2013). High-throughput classification of clinical populations from natural viewing eye movements. *Journal of Neurology*, *260*(1), 275–284. doi:10.1007/s00415-012-6631-2 PMID:22926163

Vickers, J. N. (1996). Visual control when aiming at a far target. *Journal of Experimental Psychology. Human Perception and Performance*, *22*(2), 342–354. doi:10.1037/0096-1523.22.2.342 PMID:8934848

Vincent, B. T. (2012). How do we use the past to predict the future in oculomotor search? *Vision Research*, *74*, 93–101. doi:10.1016/j.visres.2012.08.001 PMID:22917682

Voisin, S., Pinto, F., Morin-Ducote, G., Hudson, K. B., & Tourassi, G. D. (2013). Predicting diagnostic error in radiology via eye-tracking and image analytics: Preliminary investigation in mammography. *Medical Physics*, *40*(10), 101906. doi:10.1118/1.4820536 PMID:24089908

Vrzakova, H., & Bednarik, R. (2015, June). Quiet eye affects action detection from gaze more than context length. *Proceedings of the International Conference on User Modeling, Adaptation, and Personalization* (pp. 277-288). Springer International Publishing. doi:10.1007/978-3-319-20267-9_23

Wang, Y., Reimer, B., Dobres, J., & Mehler, B. (2014). The sensitivity of different methodologies for characterizing drivers gaze concentration under increased cognitive demand. *Transportation Research Part F: Traffic Psychology and Behaviour*, *26*, 227–237. doi:10.1016/j.trf.2014.08.003

Wen, Y., Zhang, X., Wang, F., & Han, J. (2015). Predicting driver lane change intent using HCRF. *Proceedings of the 2015 IEEE International Conference on Vehicular Electronics and Safety (ICVES)* (pp. 64-68). IEEE. doi:10.1109/ICVES.2015.7396895

Wen, Y., Zhang, X., Wang, F., & Han, J. (2015, November). Predicting driver lane change intent using HCRF. *Proceedings of the 2015 IEEE International Conference on Vehicular Electronics and Safety (ICVES)* (pp. 64-68). IEEE. doi:10.1109/ICVES.2015.7396895

Zank, M., & Kunz, A. (2016, March). Eye tracking for locomotion prediction in redirected walking. *Proceedings of the 2016 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 49-58). IEEE. doi:10.1109/3DUI.2016.7460030

ENDNOTES

1 see list updated by COGAIN: http://wiki.cogain.org/index.php/Eye_Trackers
2 https://www.ttl.fi/gaze2016

APPENDIX

Table 3. Review summary

Reference	Application Area / Context	Eye Tracker	Tracked Features	Approach	Results / Performance
Ajanki et al., 2009	Information retrieval (e.g., search engine queries). Predict relevant keywords currently in the user's mind.	Tobii 1750 remote	Features based on fixation sequence	Bayesian modeling	Gaze information helps to predict relevant keywords that are relevant for the user and that could be used in IR.
Ajanki et al., 2011	Mobile virtual assistant for Augmented Reality glasses with eye tracking.	Mobile experimental gaze tracker with AR display	Gaze intensity (proportion of total time on object)	Integrated system	Virtual assistant is a feasible solution
Asteriadis et al., 2008	Attention prediction while reading on a display	Self-made remote tracker	Raw gaze direction	Fuzzy neural networks	88% success rate in predicting attentiveness
Bernhard et al., 2014	Identifying target objects for gaze in 3D rendered static and dynamic stimuli on a monitor.	Tobii X50 remote	Fixations, gaze-to-object mapping	Bayesian inference, with six fixation-object mapping methods	Variable success rates between 20-95% with considerable intersubject and interscene variation
Boisvert & Bruce, 2015	Task recognition (free-viewing, object-search, saliency-viewing, explicit saliency)	Data from Koehler et al. (2014)	Fixation structure, fixated image content and scene structure	Random forest classifier	Task detection rates clearly above chance (approx chance level +20% in accuracies)
Borji et al., 2012a	Combining bottom-up and top-down models for predicting fixations in a real-world-like setup (playing video games)	IScan RK-464	Previous saccade locations, gist, and motor action related to the game (such as 2D mouse position and joystick buttons)	Hidden Markov Model (HMM)	The approach is able to predict gaze and human attention better than chance.
Borji et al., 2012b	Predicting driver's attention in computer driving games	IScan RK-464	Fixations, saccades	Integrated top-down and bottom-up influences into a linear model	Combining the features gives slightly better results than using individual features alone
Borji & Itti, 2015	Yarbus-like task prediction under differing instructions	SR Research Eyelink with chin rest	Smoothed fixation, image features (Itti model)	kNN with boosting	It is possible to detect task from the gaze tractory (+ image features)
Borji & Tanner, 2015	Comparing saliency and object-based (center-bias) visual attention.	SR Research Eyelink with chin rest	Distribution of fixations	-	Both saliency and object center-bias contributes to gaze locations at free viewing task. Model combining both to obtain better estimates of gaze trajectories proposed.
Bulling et al., 2009	Activity recognition while performing a set of typical office activity	A self-made EOG system	90 different features of eye movements	SVM	76.1% average precision, 70.5% average recall
Carrasco & Clady, 2010	Predicting reach-to-grasp intent and target object with real objects	ASL Eye-Trac 6	Saccade velocity (inverse of gaze stability)	Hidden Markov Models fusing eye tracker scene video and hand-mounted camera feed	Recognition performance between 80-90%

continued on following page

Table 3. Continued

Reference	Application Area / Context	Eye Tracker	Tracked Features	Approach	Results / Performance
George & Routray, 2015	Biometric identification using gaze trajectory	SR Research Eyelink	Fixation sequence based quantities such as fixation duration, its standard deviation, path length, skewness etc.	Radial Basis Function Network (RBF)	Claim that gaze could make a good biometric identifier, if trained over a long period of time
Greene et al., 2012	Yarbus-like task prediction under differing instructions	SR Research Eyelink 1000	Features derived from fixation sequence as well as dwell time on regions of interest		Negative result claiming that prediction cannot be performed
Hamed et al., 2016	User's preference prediction while reading keyword clouds	SMI RED 500 remote tracker	Fixation location and duration based features, pupil size	Gaussian processes	The accuracy of the best feature in the binary classification task is 63%.
Haji-Abolhassani & Clark, 2013	Yarbus-like task classification between hard and easy visual search	Iscan RK-726PCI remote tracker	Gaze points	Hidden Markov Model (HMM); Different model for easy and hard tasks.	HMM outperforms simple top-down models
Huang et al., 2015	Predicting customer selected ingredients based on gaze in salesperson-customer sandwich making scenarios	SMI Gaze tracking glasses	Fixations on food ingredients	Support vector machine (SVM)	76% accuracy in prediction 1.8s in advance of spoken request
Johnson et al., 2014	Predicting gaze behavior while driving in a simulator in three tasks: controlling speed, following a lead car, and following a lane	Not reported, integrated to the HMD	Fixations per targets, dwell times	A softmax "barrier" model integrating task importance and noise estimates to allow for uncertainty	"Similar" performance comparing KL divergence between individual human to average human and model to average distributions
Kanan et al., 2014	Yarbus-like task prediction under differing instructions	Used data collected by Greene et al. (2012)	Preprocessed features of gaze trajectory, inluding temporal information	Many	Task can be inferred by using only motor information, i.e., no information of the image by using off-the-shelf state-of-the-art classificiation algorithms. However, summary statistic alone (without time series information) may not be sufficient.
Kit & Sullivan, 2016	Everyday tasks in naturalistic environments	SMI Mobile Eye	Chronological list of discretized saccade directions and amplitudes	HMMs; maximum likelihood and maximum a posteriori for classification speed and robustness	overall performance of 36% across tasks with chance at 20%
Lethaus et al., 2011	Predicting driver's intent in a simulator	SMI iView X HED (head-mounted)	Gaze points, dwell times	Artificial Neural Networks	Left lane change is predicted better than right lane change
Lethaus et al., 2013	Predicting driver's intent in a simulator; how early can intention be predicted?	SMI iView X HED (head-mounted)	Gaze points, dwell times	Artificial Neural Networks	Above change prediction up to 5 seconds before event

continued on following page

Table 3. Continued

Reference	Application Area / Context	Eye Tracker	Tracked Features	Approach	Results / Performance
Lethaus et al., 2013 (Neurocomputing)	Predicting driver's intent in a simulator; Which model works best and how well?	SMI iView X HED (head-mounted)	Gaze points, dwell times	Artificial Neural Networks, Bayesian Networks, and Naive Bayes Classifiers	ANN seems to be the best predictor but with a small difference
Liut et al. 2009	Predict differences in skill-level (novices vs. experts) while reading and building concept maps	Tobii 1750	Gaze location and fixation durations	Hidden Markov Models	96% accuracy in differentiating novices from experts
Marius t'Hart et al., 2012	Path navigation in real-world city environments	EyeSeeCam	Eye-in-head and gaze-in-world coordinates	Rough classification of gaze location	comparisons of basic eye movement metrics, statistics and distributions between different conditions
Mathe & Sminchisescu, 2015	Viewing short video clips from Hollywood movies and various sports	SMI iView X Hispeed, 500Hz	Fixations	Dynamic saliency for video content using dynamic histogram-of-gradient and motion boundary histograms	training saliency predictors based on gaze data; two annotated action recognition datasets for gaze data supplied
Oertel et al., 2011	Involvement in spontaneous conversation, how to predict using gaze, blinks, audio cues	None	Blinks and whether a person looked at the conversation partner or not	Use standard SVM (radial basis function) to estimate involvement using covariates such as case, blinks, audio cues.	Accuracy of prediction 68%, gaze seems to correlate with involvement.
Peng et al., 2015	Predicting lane changing while driving in real traffic	faceLAB 5	Gaze locations on predefined targets (windshield, dashboard, rearview mirror)	Back-propagation neural network model	Prediction accuracy was 85,4% 1,5s before lane change
Puolamäki et al., 2005	Inferring (word) relevance in information retrieval	Tobii 1750	Features computed from fixation sequence	Hidden Markov Models	Information extracted from gaze can be used to aid in information retrieval task when combined with contextual information
Puolamäki et al., 2008	Gaze-based proactive information retrieval; supporting information finding while browsing hypertext	Tobii 1750	19 eye movement features: number of, duration, of fixations etc.	Applying term-specific eye movement patterns to a SVM based document search	Gaze-enhanced method outperforms baseline BM25 ranking method (p=.047), although performance is modulated by search task
Rothkopf, 2016	Walking in VR	Applied Science Laboratories 501	Sequence of gaze locations converted to a codebook	Modeling codebook sequences of gaze with HMMs with variable number of latent variables	The presented models generate similar gaze sequences to human observers.
Vincent, 2012	Understanding mechanisms in visual search while searching for targets on a display	SR Research Eyelink 1000 remote tracker	Gaze location	Different models of search mechanisms	Best accuracy with the model assuming learning 2nd order statistics and that world is dynamic

continued on following page

Table 3. Continued

Reference	Application Area / Context	Eye Tracker	Tracked Features	Approach	Results / Performance
Voisin et al., 2013	Predicting diagnostic errors in mammography analysis based on radiologists' gaze behavior on a laptop screen and image characteristics	Mirametrix S2	ROI-based eye movement and pupil dilation features	Genetically selecting best performing machine learning algorithms per subject/subject group and feature set	Initial results (limited by number of cases and participants) showing that machine learning methods can be applied to predicting human error in diagnostic scenarios
Vrzakova & Bednarik, 2015	Organizing on-screen content using a mouse in a problem-solving task	Tobii 1750	54 gaze features per gaze sequence	SVM with an radial-basis-function kernel	While increasing fixation sequence length before action improves intent recognition, including the "quiet eye" fixation just before action initiation outperforms length optimization by approx. 15%
Wen et al., 2015	Predicting lane changing in simulator driving	SMI RED 500	Gaze x position, and it's derivative	Hidden Conditional Random Fields (HCRF)	Prediction accuracy was 99% 0.5s before lane change, 85% 2.0s before.
Zank & Kunz, 2016	Predicting user locomotion to alleviate redirected walking in 3D virtual environment	SMI eye tracker integrated in an Oculus DK2 HMD	Gaze points	Bayesian model for locomotion target and gaze point / location	Improves the prediction to some extent, as compared to approach without utilizing gaze

Kristian Lukander, MSc (Tech), is a Research Engineer at the Finnish Institute of Occupational Health. He is currently working on his PhD on gaze tracking in naturalistic environments. His work has concentrated on methodological development for visualizations, user-centered research and psychophysiological methods in applied occupational settings.

Miika Toivanen made his doctoral theses in Aalto University, Finland, about computer vision and computational methods. Recently, he has been developing gaze tracking algorithms and mobile devices in FIOH. Toivanen is currently working in University of Helsinki where he develops and uses mobile gaze tracking technology to measure student's learning behavior in classroom. His interests are in applied mathematics and programming.

Kai Puolamäki is Senior Research Scientist at the Finnish Institute of Occupational Health. He completed his PhD in 2001 in theoretical physics at the University of Helsinki. His primary interests lie in the areas of computational data analysis, data mining, machine learning, and related algorithms. One of the substance areas of his interest is modelling of gaze trajectory. Dr Puolamäki holds a title of docent in information and computer science at the School of Science of Aalto University, Finland. Dr Puolamäki has a web site at http://www.iki.fi/kaip/.